AD-A262 702

DTIC
S ELECTE
c APR 9 1993 D

ARMSTRONG LABORATORY

ARMSTRONG LABORATORY

EXPLICIT, IMPLICIT, AND SUBJECTIVE RATING MEASURES OF SITUATION AWARENESS IN A MONITORING TASK (U)

Martin L. Fracker, Major, USAF

CREW SYSTEMS DIRECTORATE
HUMAN ENGINEERING DIVISION

Sharon A. Davis

LOGICON TECHNICAL SERVICES, INC.
DAYTON, OHIO

93-07416

OCTOBER 1991

FINAL REPORT FOR PERIOD JANUARY 1990 TO JANUARY 1991

93 4 08 058

AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433-6573

NOTICES

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Please do not request copies of this report from the Armstrong Aerospace Medical Research Laboratory. Additional copies may be purchased from:

<div style="text-align:center;">

National Technical Information Service
5285 Royal Road
Springfield, Virginia 22161

</div>

Federal Government agencies and their contractors registered with the Defense Technical Information Center should direct requests for copies of this report to:

<div style="text-align:center;">

Defense Technical Information Center
Cameron Station
Alexandria, Virginia 22314

</div>

TECHNICAL REVIEW AND APPROVAL

<div style="text-align:center;">

AL-TR-1991-0091

</div>

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

The voluntary informed consent of the subjects used in this research was obtained as required by Air Force Regulation 169-3.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

KENNETH R. BOFF, Chief
Human Engineering Division
Armstrong Laboratory

INTRODUCTION

Situation awareness (SA), or what operators know about their immediate tactical situation, has typically been studied in the context of operational military tasks such as aerial combat (Endsley, 1989) or strategic bombing (Marshak et al., 1987). Generalization of research results to operational military environments is clearly an advantage of such tasks. But the disadvantage is some uncertainty as to the source of the observed effects. A recent example of this uncertainty may be found in Fracker (1991a), who evaluated several candidate measures of SA in a game-like task incorporating various elements of a military mission. Fracker found that subjects' ability to correctly show the locations of enemy assets was unaffected by the uncertainty as to whether those assets actually belonged to the enemy. This somewhat counterintuitive result could have been caused by some insensitivity inherent in the location performance measure or by some masking effect produced by the complex nature of the subjects' task. This ambiguity suggests that there may be value in also evaluating candidate SA measures in simpler, more abstract laboratory tasks where greater experimental control is possible. Measures that continue to exhibit unwanted characteristics in well-controlled laboratory settings could then be deemed a low priority for further development.

The present experiment was designed to replicate the essential features of Fracker's (1991a) experiment 1 using an abstract monitoring task rather than the combat task of the earlier research. In that first experiment, Fracker had subjects engage in combat with enemy "aircraft" represented by uniquely shaped objects that were red in color. Subjects controlled a blue object which they used to destroy red objects before the red objects could destroy them. In addition, there were also other friendly and neutral (gray) objects present which subjects were to refrain from destroying.

Fracker (1991a) evaluated two classes of SA measures, explicit and implicit (see Fracker, 1991b, for a discussion of these and other classes). In the explicit class, subjects were tested on the location or the identity of specific objects. Thus, the combat task was periodically frozen, one of the objects was removed from its correct location, and its color was removed. Subjects simply moved the object back to its correct location or else indicated the object's color. Location error (measured in degrees of visual angle) was the variable of interest for the location probe; in the color probe, the relevant measures were of reaction time and accuracy. The implicit class included only one measure, an estimate of subjects' ability to discriminate between when enemy objects were and were not within range of the subjects' weapon (referred to as "envelope sensitivity" and calculated as A'; see Macmillan and Creelman, 1990).

In his experiment, Fracker (1991a) manipulated combat intensity by changing the number of enemy aircraft (the total number of aircraft

1

remained constant--when there were more enemies, there were fewer
neutrals). This manipulation led to larger location errors, slower
and less accurate color probe responses, and poorer envelope
sensitivity. Fracker also manipulated the difficulty of keeping track
of objects' identities as friend, foe, and neutral. In one condition,
objects retained their identity throughout a trial; in another,
objects randomly switched identities several times during a trial.
Identity inconsistency proved disruptive in the color task, leading to
less accurate (but not slower) responses, but had no effect whatsoever
on location error or envelope sensitivity.

These results are not easy to explain. On one hand, resource
theory (Kahneman, 1973; Wickens, 1980) can explain why increased
combat intensity proved generally detrimental across tasks (combat
task, location probe, color probe). On the other hand, the failure of
identity inconsistency to affect the location or combat tasks is
problematic for resource theory. The theory could account for the
observed difficulty insensitivity in the location task by supposing
that identity processing uses a resource not used in location
processing, a possibility consistent with Wickens' (1980, 1984) views
on multiple resources. But the theory can not easily account for the
insensitivity observed in envelope sensitivity because identity
processing must necessarily be involved in discriminating enemies from
non-enemies. In experiments 2 and 3, Fracker (1991a) explored this
question by directly evaluating the multiple resource hypothesis:
subjects performed the combat task simultaneously with either a
spatial or verbal processing task. Contrary to multiple resource
theory, these tasks interfered with each other (experiment 3);
contrary to single resource theory, these tasks did not interfere with
the combat, location, or color tasks.

Fracker's (1991a) results might lead to a rejection of resource
theory in favor of one of its more recent competitors (e.g., Hirst and
Kalmar, 1987; Navon and Miller, 1987), but only if one is sure that
the location and sensitivity measures themselves were not at fault.
Regarding location error, Fracker observed that the measure was
unstable and suggested that subjects may have been responding to the
location probes with less precision than the measure assumed.
Fracker suggested constraining location responses so as to bring about
a better match between assumed and actual precision. At the same time,
envelope sensitivity, as it was measured in Fracker's study, may have
been more readily influenced by location than by identification
processes.

The Present Experiment

In the present experiment, subjects performed a monitoring task
in which six objects moved across a grid by jumping from node to node
in a predictable pattern. This grid was intended to reduce the
possible locations of the objects to the number of nodes and to

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE July 1991	3. REPORT TYPE AND DATES COVERED Interim, Oct 1990 to July 1991

4. TITLE AND SUBTITLE Explicit, Implicit, and Subjective Rating Measures of Situation Awareness in a Monitoring Task	5. FUNDING NUMBERS F33615-89-C-0532 PE 62202F PR 7184 TA 14 WU 25
6. AUTHOR(S) Martin L. Fracker, Major Sharon A. Davis *	

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Armstrong Laboratory Crew Systems Directorate, Human Engineering Division, AFSC, HSD Wright-Patterson AFB OH 45433-6573	8. PERFORMING ORGANIZATION REPORT NUMBER AL-TR-1991-0091

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
*Logicon Technical Services, Inc.
P.O. Box 317258, Dayton OH 45431-7258

12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.	12b. DISTRIBUTION CODE A

13. ABSTRACT (Maximum 200 words)   The situation awareness (SA) and mental workload of 56 subjects were evaluated as they monitored one or more attributes of six objects moving systematically over a rectangular grid. Subjects were assigned to one of seven groups depending upon whether they were to monitor object locations (location task), object colors (color task), whether the objects flashed (flash task), or some combination of these three. Both task performance and subjective ratings were used to assess subjects' awareness of the three object attributes. In addition, subjective ratings of mental workload were collected. All subjects performed the monitoring task under four different conditions formed from the factorial combination of 1) the probability that objects of a certain color would flash and 2) whether object colors remained consistent or changed during the course of a trial. The results pointed to the usefulness of both flash and color task performance as measures of SA. Subjects were very poor at the location task, suggesting either their location awareness was poor or the location task is not a good measure of that awareness. Subjective ratings proved useful but occasionally dissociated from task performance. One possibility is that subjective ratings reflect rational inferences by the subjects rather than the outcome of their introspections.

14. SUBJECT TERMS Situation Awareness Mental Workload Human Performance Subjective Measurement		15. NUMBER OF PAGES 32
		16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UNLIMITED

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to *stay within the lines* to meet *optical scanning requirements*.

Block 1. Agency Use Only (Leave blank).

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	-	Contract	PR	-	Project
G	-	Grant	TA	-	Task
PE	-	Program	WU	-	Work Unit
		Element			Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

> DOD - See DoDD 5230.24, "Distribution Statements on Technical Documents."
> DOE - See authorities.
> NASA - See Handbook NHB 2200.2.
> NTIS - Leave blank.

Block 12b. Distribution Code

> DOD - Leave blank.
> DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
> NASA - Leave blank.
> NTIS - Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

PREFACE

DTIC QUA~~

Accesion For		
NTIS CRA&I		☒
DTIC TAB		☐
Unannounced		☐
Justification		
By		
Distribution /		
Availability Codes		
Dist	Avail and/or Special	
A-1		

TABLE OF CONTENTS

LIST OF TABLES

constrain where subjects could place the objects during the location
task. Thus, the grid was expected to effect a better match between
the precision of the subjects' location responses and the precision
with which the accuracy of those responses was measured.

The objects were the same as those used in Fracker's (1991a)
experiments. Two of these objects were red, two were blue, and two
were green. Occasionally, one of the objects would flash. In the
flash task, subjects were required to detect an object's flash and to
indicate the flashing object's color. Because color identification
was an integral part of the flash detection task, color inconsistency
(manipulated in the same way as in Fracker, 1991a) should have
affected A', or the ability to discriminate flashes from non-flashes,
by causing sensitivity to decrease.

As in Fracker (1991a), the location and color tasks occurred
during freezes of the monitoring task. Fracker had reported that
location task performance was more accurate for enemy objects rather
than for either friendly or neutral objects. A likely explanation for
this greater accuracy is that subjects allocated more attention to
enemies than to others. An alternative explanation is that subjects
knew their own location and could guess that enemies would normally be
in the same vicinity, a guess that they could not make for friendlies
or neutrals. The present experiment attempted to avoid this ambiguity
by manipulating attention allocation directly. On some trials, red
objects were as likely to flash as either blue or green objects; on
other trials, red objects were more likely to flash than either blue
or green objects. Subjects performing the flash task were expected to
allocate more attention to objects based on their flash probability;
thus, the accuracies of their location and color task responses were
also expected to be influenced by flash probability.

Unlike Fracker's experiments, in which all subjects performed
all three tasks, subjects in the present experiment were divided into
seven groups. Three of the groups performed just one of the three
tasks, three of the groups performed a pair of tasks, and the seventh
group performed all three. Thus, the present experiment was able to
assess the effect of each task on the other two.

In terms of a single resource model, anything that makes one task
more difficult may influence performance on other concurrent tasks
depending upon the subjects' allocation strategies. For example, if
the color identification task becomes more difficult, subjects might
maintain the quality of their color task performance but only at the
cost of poorer location task performance. On the other hand, subjects
might choose to let color task performance degrade while maintaining
performance on the location task. If the flash task is included with
the location and color tasks, then the subjects can make even more
complex allocation decisions.

3

Knowing what allocation choices that subjects have made is clearly going to influence whether one interprets their performances on the various tasks as consistent with resource theory. One method for attaining such knowledge is to attempt to manipulate such task allocation choices experimentally, an approach that has been severely criticized by Navon (1984, 1985, 1990). Another method is to infer allocation strategies from observed performances--which obviously becomes circular when those inferred strategies are then used to interpret performance. A third approach--adopted in the present experiment--is to allow subjects to spontaneously adopt their own allocation strategies and then to assess those strategies independently of task performance.

In order to assess subjects' attention allocation strategies, a paired-comparison subjective rating approach was used. In order to assess which tasks received the most attention, subjects were presented with three pairs of object attributes, one pair at a time: location-color, location-flash, color-flash. For each pair, subjects provided two different sets of ratings. In one set, subjects indicated which attribute was more critical and rated how more critical that attribute was compared to the other. In the second set, subjects indicated to which attribute they spent the most time attending and rated how much more time that attribute received compared to the other. In a similar fashion, subjects also rated the time they spent attending to objects by color. These paired-comparisons were then used to obtain scale values for each object attribute (or color) using techniques described by Budescu, Zwick, and Rapoport (1986; see also Saaty, 1977; Vidulich, 1989). These scale values were interpreted as indices of the relative amount of attention allocated to each attribute or color. In the case of attributes, the scale values then also indicated the subjects' allocation strategy across tasks.

Finally, the present experiment supplemented the main explicit and implicit measures of SA with subjective rating measures, again using the paired-comparisons approach just described (see Fracker, 1991b; Fracker and Davis. 1990). The same approach was also used to assess subjective mental workload (see Vidulich, 1989).

METHOD

Subjects

Fifty-six paid volunteers from the Wright State University Community served as subjects. All subjects had normal or corrected-to-normal vision.

Task Overview

Subjects monitored six colored, uniquely-shaped objects moving around on a white grid against a black background (see Figure 1).
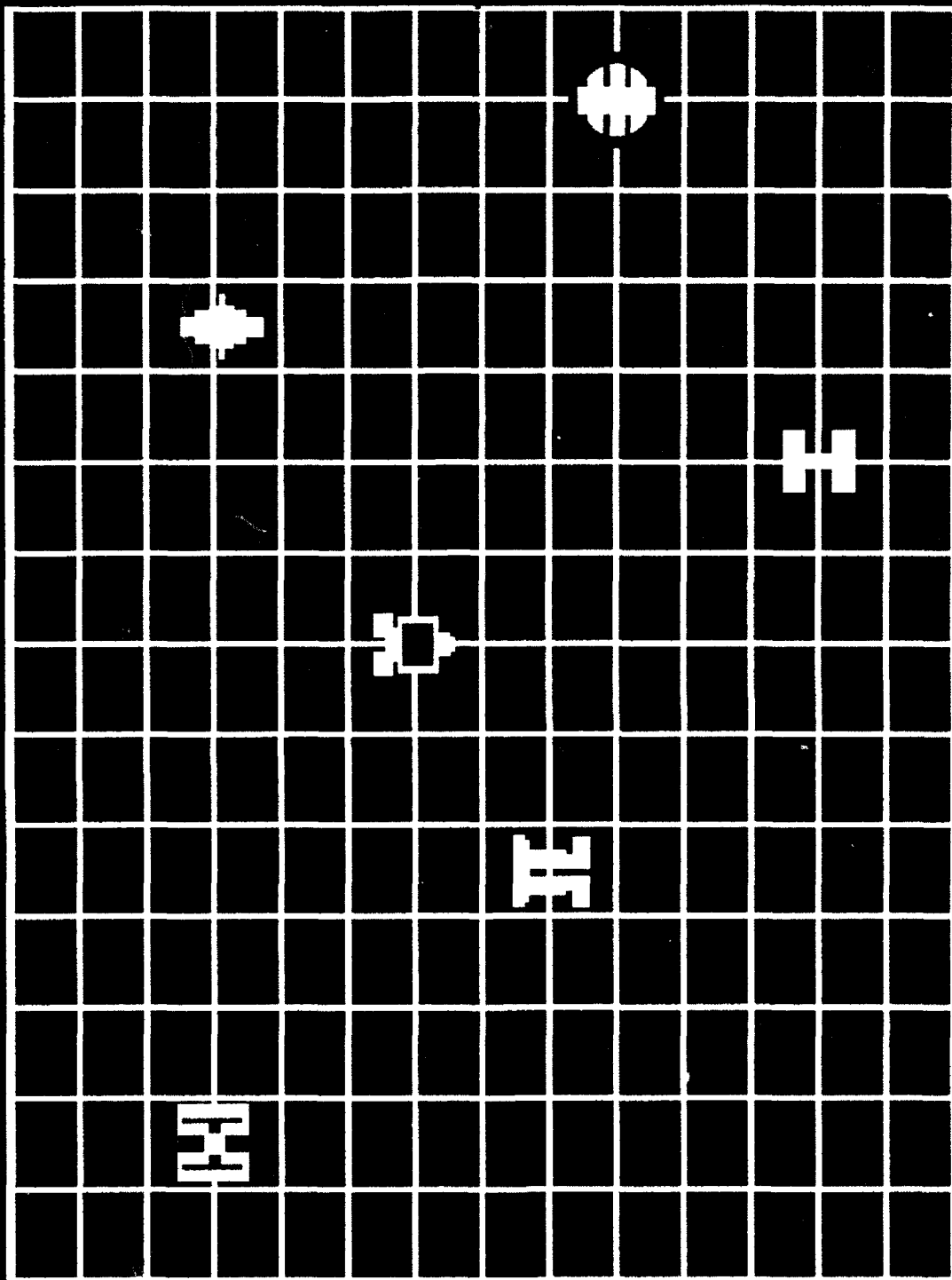
4

Figure 1. Grid Viewed by Subjects.

The six objects varied slightly in size, ranging from .5 cm X .9 cm to 1 cm X 1 cm. With a viewing distance of 55.5 cm, each cm corresponded to approximately 1 deg of visual angle. The outline of the grid was a basic white square measuring 14 cm X 18.5 cm. The individual cells comprising the grid were formed by 15 horizontal and 15 vertical white lines. The interior of the grid was thus comprised of 196 individual cells, each measuring 1 cm X 1.33 cm each. Below the grid was a box measuring 3.5 cm X 2 cm which was used for the color identification and location tasks (described below).

Objects moved across the grid by "stepping" from one intersection, or "node," to the next in a completely predictable manner. Three of the six objects moved horizontally and three moved vertically. All objects started either from the top or the right side of the grid and moved either towards the left side or the bottom of the grid, one node at a time, at a speed of one node every 227 ms. Once an object reached the bottom (or side) of the grid, it moved to an adjacent node and began traversing back up (or across) the grid. Thus, horizontally moving objects gradually worked their way down and back up the grid, and vertically moving objects gradually worked their way across the grid one direction and then the other.

Objects were assigned one of three colors randomly at the start of each new trial: red, blue, or light green. For half the trials, this assignment remained constant for the duration of the trial. For the other half of the trials, object colors were randomly reassigned 12 times during the trial. These color reassignments simply occurred randomly every 15 to 45 s with an average assignment duration of 30 s. When either the location or color tasks were performed, these random reassignments occurred in the middle of an inter-freeze interval (see below).

Subjects were assigned to one of seven different experimental groups depending upon which task or tasks they would perform. There were three single task groups, three dual task groups, and one triple task group, representing all possible combinations of three tasks: flash detection, color task, and location task.

Task Descriptions

Flash Detection Task

At random intervals during a trial, one of the six objects would flash. That is, the object would briefly change from its normal color (red, blue, or light green) to a dark gray; 250 ms later, the object returned to its original color. Dark gray was chosen as the flash color to reduce the chance that the flash would attract subjects' attention if they were not attending to the object. Subjects were to press a single button with their left index finger in order to indicate that they had detected the flash. A button press would

interrupt the trial regardless of whether an object had actually flashed, and the software then prompted the subject for the flashing object's color. Subjects responded by pressing one of three color buttons with their right hand.

Each trial lasted 360 s, and 45 flashes occurred during this period. On average, flashes occurred 8 s apart, although inter-flash intervals actually varied randomly from 4 to 12 s in duration. On half the trials, red, blue, and green objects were equally likely to flash ($p = 1/3$). On the other half, red objects were more likely to flash ($p = 1/2$) and green objects less likely to flash ($p = 1/6$). Selection of which object flashed was random with replacement, constrained only by the active flash probabilities.

Following each object flash, the subject had 2 s in which to press the flash detection button. Failures to respond within 2 s were recorded as "misses." If detected, subjects had unlimited time in which to indicate the flashing object's color. Following the color response, the computer provided feedback on their selection accuracy (correct or incorrect). The software maintained a count of correctly identified flashes (hits), flashing objects not noticed (misses), and false alarms by object color. Correct rejections were defined as an iteration of the main program loop in which no detection response occurred and for which no flash had occurred within the last two seconds. At the end of a trial, the subject was provided with feedback on the number of hits, misses, false alarms, and average reaction time for hits over the duration of the trial.

Color Identification Task

At 12 randomly selected times during a trial, the objects "froze" (i.e., all movement completely stopped) and all the objects turned white. The intervals between these freezes varied in duration from 15 to 45 s but averaged 30 s in length. At the moment of the freeze, three objects disappeared from the grid and reappeared one at a time in the box at the bottom of the screen. Objects to be tested were chosen randomly with the constraint that, over the course of the trial, each object was tested six times.

The subject's task was to identify the color which the object had at the moment of the freeze by pressing one of the three color identification keys. Subjects had 60 s in which to press a key, after which the software recorded an incorrect response. Feedback as to whether the response was correct immediately followed, the object's color was restored, and the object was returned to its correct grid location. Two seconds later, the next of the three objects to be tested appeared in the box. For each object, the software recorded the speed and accuracy of the subjects' response.

Location Task

Like the color identification task, the location task was performed during each of the 12 freezes. When subjects performed both the location and color tasks, both were performed during the same freeze. Half the subjects performed the location task first, and half performed the color task first.

Again, three objects disappeared from the grid and reappeared one at a time in the box at the bottom of the screen. And again, objects to be tested were chosen randomly with the constraint that, over the course of the trial, each object was tested six times. Subjects were to return the object to its correct location on the grid by pointing to and touching the grid at the desired location. When the subject touched the grid, the object immediately moved from the box to grid intersection (referred to as a "node") closest to the point which the subjects had touched. The software then moved the object to its correct location, thus providing visual feedback as to the subjects' accuracy. As with the color identification task, there was a two second delay before the next object to be tested appeared in the box. The software initially calculated subjects' location error as the Euclidean distance in pixels between the subject's placement of the object and the object's actual location; this distance was then converted to degrees of visual angle and recorded. At the conclusion of a trial, average location error was calculated and displayed to the subject.

Subjective Ratings

Overview

Three sets of ratings were collected following each trial of the second session: criticality of object attributes (color, location, and whether the object flashed), time-spent-attending to object attributes, and time-spent-attending to objects by color (red, blue, and light green). "Criticality" was defined as the importance to successful task performance. In addition, four sets of ratings were collected at the end of the second session: in these ratings, subjects rated the four within-subject experimental conditions in terms of their awareness of object location, awareness of object color, awareness of whether objects flashed, and mental workload.

Using a paired-comparison scaling technique, all ratings were collected in exactly the same way. Each pair of stimuli to be compared were presented at opposite ends of a box measuring 20.3 cm long and .9 cm high comprised of 19 equally-spaced increments. The center of the box was marked "EQUAL." The two stimuli for a given comparison appeared just below the box, one stimulus at either end. Centered below the word "EQUAL" was the name of the dimension on which the stimuli were to be compared. Subjects indicated which stimulus was

higher on the named dimension by drawing a line from the center of the box in the direction of the selected stimulus. The length of the line drawn indicated how much greater the selected stimulus was compared to the other on the named dimension. Subjects used the left and right cursor keys on the computer keyboard in order to draw the line. By default, one press of a cursor key would draw the line to the next increment. However, subjects could change this feature (by pressing the down arrow key) in order to allow them to draw lines less than one increment at a time.

Once subjects had drawn the line in the desired direction and to the desired distance, they pressed the keyboard enter key. At this time, the software computed a ratio for one stimulus to the other. A line of zero length represented a ratio of 1:1. A line one increment in length represented a ratio of 2:1, a line of two increments represented a ratio of 3:1, and so on. Once all possible pairs among the stimuli had been rated in this manner, the software computed scale values for the stimuli as described by Budescu, Zwick, and Rapoport (1986). These scale values then became the ratings analyzed in the present report.

Training

Prior to the beginning of the second session, subjects were trained to use the paired-comparison software. In order to ensure that they understood the rating concept and procedure, subjects were asked to rate all possible pairs of five numerical stimuli (100, 300, 500, 700, and 900) on the dimension of "quantity." Subjects were then able to view the scale values which their ratings had produced. If subjects failed to scale the numbers in a strictly increasing monotonic fashion, then the instructions were repeated and subjects were asked to re-scale the numbers.

Trial Ratings

Following each trial, subjects performed three sets of ratings. In one set, the dimension name "Criticality" appeared below the rating scale. All possible pairs of the three attribute names "Location," "Color," and "Flash" then appeared, one pair at a time. In a second set subjects rated the time spent attending to those same attributes. Thus, the dimension name was "Time Spent Attending," and the stimuli were again "Location," "Color," and "Flash." Finally, subjects also rated the time spent attending to objects by color. The dimension name was again "Time Spent Attending," and the stimuli were "RED Objects," "BLUE Objects," and "GREEN Objects."

9

Session Ratings

At the end of the session, subjects compared the four within-subject experimental conditions on four dimensions. The stimuli identifying the four conditions were:

```
Flash EQUAL : Color STAYS_SAME
Flash EQUAL : Color CHANGES
Flash _RED_ : Color STAYS_SAME
Flash _RED_ : Color CHANGES
```

First, subjects compared the conditions in terms of the mental workload experienced while performing their tasks. For these comparisons, the dimension name "WORKLOAD" appeared below the rating scale. Next, subjects compared the experimental conditions on three dimensions of SA, one dimension at a time. Each dimension corresponded to one of the three object attributes: location, color, and whether the objects flashed. Thus, the dimension names for these three ratings were "AWARENESS of Object LOCATIONS," "AWARENESS of Object COLORS," and "AWARENESS of FLASHING Objects," respectively.

Apparatus

The experimental conditions were generated by a Zenith 248 (IBM AT-compatible) personal computer and displayed to the subject via a Zenith composite color monitor located inside a sound-attenuated booth. A TSD Display Products, Inc. transparent touchscreen was mounted to the front of the monitor, and was used by the subjects to respond to the location probes. Subjects sat in a chair with one custom-made response box mounted onto each arm: a one-button box mounted on the left arm, and a three-button box mounted on the right. Subjects placed their chins in custom-made chin-rest which fixed their viewing distance from the monitor at 55.5 cm.

Subjective ratings were collected by a Zenith lap-top personal computer (IBM XT-compatible) which sat on a small platform located on the subject's left.

Sessions

All subjects participated in two experimental sessions. Each subject was randomly assigned to one of the seven experimental groups at the start of the first session. Session one was a practice day only, and was intended to familiarize subjects with their assigned tasks and with each of the four within-subject experimental conditions. Actual data collection was conducted during the second session only. The order in which subjects received the four experimental conditions was counterbalanced using two 4 X 4 randomized Latin squares for the eight subjects in each group. Thus, order effects were counterbalanced within every four subjects.

10

RESULTS

The results are reported in three sections: Subjective Time and Criticality Ratings, Subjective SA and Workload Ratings, and Task Performance Measures. The general approach in each section was to analyze related ratings or measures together in a single 7 x 2 x 2 mixed MANOVA (Group by Flash Probability by Color Inconsistency) where group was a between-subjects variable while flash probability and color inconsistency were within-subjects variables. In order to guard against an escalating Type 1 error rate, individual ratings or measures were examined using univariate ANOVA's only if the parent MANOVA led to rejection of the null hypothesis.

Two errors in collecting subject data led to unequal numbers of subjects in the seven groups. First, one subject was inadvertently run in the wrong group for some of the trials, but this fact was not discovered until the data were under analysis. This subject was removed from the sample. For analysis of the task performance measures, one subject from each of the remaining six groups was also deleted at random in order to preserve equal n's across groups and permit an orthogonal analysis of variance. Second, all subjective rating data from five subjects were inexplicably lost. This error--in addition to the previously deleted subject--reduced the sample size to seven in two groups, and to six in two others. In order to preserve statistical power, all subjective rating data were analyzed for the remaining 50 subjects using the general linear model approach to the nonorthogonal analysis of variance.

Subjective Time and Criticality Ratings

Group effects. Table 1 displays each group's subjective time and criticality ratings for object attributes (location, color, flash). Manipulation of flash probability did not affect these ratings (p's > .14), so ratings for only the consistent and variable color conditions are shown.

11

Table 1. *Subjective Time and Criticality Ratings for Object Attributes.*

		Criticality of			Time Spent Attending To		
		Location	Color	Flash	Location	Color	Flash
L:	Con[1]	.67	.15	.18	.71	.22	.08
	Inc	.68	.19	.13	.76	.17	.08
C:	Con	.14	.74	.11	.18	.69	.13
	Inc	.11	.77	.12	.11	.77	.12
F:	Con	.28	.30	.42	.31	.25	.43
	Inc	.30	.33	.37	.34	.27	.39
LC:	Con	.68	.26	.06	.64	.31	.05
	Inc	.46	.48	.06	.35	.60	.05
LF:	Con	.51	.18	.32	.42	.17	.42
	Inc	.42	.21	.38	.40	.22	.39
CF:	Con	.13	.38	.49	.12	.30	.57
	Inc	.12	.45	.43	.10	.37	.53
LCF:	Con	.33	.32	.35	.31	.26	.42
	Inc	.20	.44	.36	.14	.46	.40

[1] L = Location Task, C = Color Task, F = Flash Detection Task.
   Con = Colors stay consistent during a trial.
   Inc = Colors change (are inconsistent) during a trial.

Task group did not influence the ratings for colors ($p > .8$) but did affect the ratings for object attributes, both criticality (*Wilks' Lambda* = 0.093, $F(18,116) = 8.49$, $p < .0001$) and time-spent-attending (*Wilks' Lambda* = 0.07, $F(18,116) = 10.09$, $p < .0001$). Univariate ANOVAs for the criticality ratings showed that group differences existed for all three object attributes (Location: $F(6,43) = 13.45$, $p < .0001$; Color: $F(6,43) = 25.69$, $p < .0001$; Flash: $F(6,43) = 8.08$, $p < .0001$). The protected least significance difference approach (the Fisher test; see Keppel, 1982) was used to examine the differences among group means. With the exception of the triple task group, any group that performed the location task assigned location a higher criticality than any group that did not perform the location task (p's $< .01$). Similarly, groups that performed the color task rated color higher in criticality than groups not performing the color task (p's $< .01$). Further, a "dilution effect" appeared in the ratings of color: as tasks were added to the color task, the criticality assigned to the color attribute decreased (p's $< .01$). A similar dilution effect had

also appeared in the ratings of location but was weaker and less consistent. Finally, groups that performed the flash task rated the flash higher in criticality than groups not performing the flash task (p's < .01); there was no evidence of a dilution effect for flash ratings, however.

Univariate ANOVAs for the time-spent-attending-to-attributes ratings closely paralleled those for criticality (Location: $F(6,43)$ = 19.76, p < .0001; Color: $F(6,43)$ = 19.90, p < .0001; Flash: $F(6,43)$ = 16.69, p < .0001). The major difference compared to the criticality ratings was that the dilution effect in the time-spent ratings for location was consistent and reliable (p's < .01). Otherwise, the differences among groups were virtually identical in the two sets of ratings.

Color inconsistency effects. Color inconsistency effects appeared for both the criticality and time-spent ratings for object attributes (Criticality: *Wilks' Lambda* = 0.68, $F(3,41)$ = 6.40, p < .002; Time-spent: *Wilks' Lambda* = 0.62, $F(3,41)$ = 8.55, p < .0002). Univariate ANOVAs again showed that only criticality ratings of location and color were effected by color inconsistency (Location: $F(1,43)$ = 14.08, p < .0005; Color: $F(1,43)$ = 13.60, p < .0006; Flash: p > .5). Compared to color consistency, color inconsistency led to a decrease in the criticality of location (.38 to .32) but an increase in the criticality of color (.34 to .41). Precisely the same pattern appeared in the time-spent ratings, where color inconsistency led to less time attending to location (.37 to .32; $F(1,43)$ = 15.25, p < .0003) and more time spent attending to color (.32 to .40; $F(1,43)$ = 24.24, p < .0001).

The effect on color ratings was only marginally reliable (see Table 3), *Wilks' Lambda* = 0.83, $F(3,41)$ = 2.75, p < .06. Univariate ANOVAs revealed marginal effects only for red and blue objects (Red: $F(1,43)$ = 3.85, p < .06; Blue: $F(1,43)$ = 4.35, p < .05; Green: p > .8). Color inconsistency led to more time attending to red objects (.39 to .42) and less time attending to blue objects (.34 to .31).

Group by color inconsistency interactions occurred in both the attribute and colors time-spent ratings (Attributes: *Wilks' Lambda* = 0.41, $F(18,116)$ = 241, p < .003; Colors: *Wilks' Lambda* = 0.45, $F(18,116)$ = 2.13, p < .009. In the attributes ratings, the interaction occurred in both the location and color ratings (Location: $F(6,43)$ = 6.69, p < .0001; Color: $F(6,43)$ = 5.62, p < .0002; Flash: p > .9). These interactions in the attribute ratings resulted because the color inconsistency effect occurred only in those groups performing *both* the location and color tasks (p < .01). In the time-spent-attending-to-*colors* ratings, the interaction occurred only in the ratings for green objects (see Table 2), $F(6,43)$ = 4.28, p < .002 (Red: p > .27; Blue: p > .09). This interaction resulted because the color inconsistency effect occurred only for the group performing

just the flash and color tasks: compared to the consistent colors condition, color inconsistency led to more time attending to green objects (.15 to .28).

Table 2. Ratings of Time Spent Attending to Green Objects By Group

Group	Consistent Colors	Inconsistent Colors	Difference
L	.34	.27	.07
C	.27	.21	.06
F	.33	.28	.05
LC	.30	.31	-.01
LF	.23	.23	.00
CF	.15	.28	-.13 ($p < .01$)
LCF	.26	.27	-.01

Flash probability effects. Flash probability had no main or interaction effects in the criticality or time-spent-attending-to-attributes ratings (p's > .1). In the ratings of time spent attending to different colors, flash probability was expected to produce a group by probability interaction: different flash probabilities were expected to alter the subjects' biases to attend to red, blue, and green objects--but only for subjects who performed the flash detection task. This interaction did not occur, however (p > .7). Instead, an overall main effect of flash probability appeared (see Table 3), *Wilks' Lambda* = .80, $F(3,41) = 3.43$, $p < .03$. Subsequent univariate ANOVAs showed that the effect occurred mainly for red objects, $F(1,43) = 10.76$, $p < .003$. A marginal effect also appeared for blue objects, $F(1,43) = 3.66$, $p < .07$. The effect was clearly unreliable for green objects (p > .15). Compared to unbiased (equal) probabilities, a red bias led to more time attending to red objects (.38 to .43) and slightly less time attending to blue objects (.34 to .31).

Table 3. Color Inconsistency and Flash Probability Effects on Ratings
of Time Spent Attending to Color.

Flash Probability	Colors	Time Spent Attending To		
		Red	Blue	Green
Equal	Consistent	.36	.37	.27
Equal	Inconsistent	.41	.31	.28
Red	Consistent	.42	.31	.26
Red	Inconsistent	.44	.31	.25

Subjective SA and Workload Ratings

Subjects' comparisons of their SA and workload across
experimental conditions are shown in Table 4.

Group effects. No main effect of group was possible in this
analysis because the ratings across within-subject conditions were
scaled to a mean of .25 (i.e., 1.0 divided by the number of
conditions). As a result, all seven groups were guaranteed to have
the same mean, and they did ($p > .9$).

Color inconsistency effects. A main effect of color
inconsistency is evident in Table 4 and was reliable, *Wilks' Lambda* =
0.33, $F(4,40) = 20.28$, $p < .0001$. Univariate ANOVAs showed that the
effect was reliable in all of the ratings except those for color
(Location: $F(1,43) = 32.56$, $p < .0001$; Color: $p > .18$; Flash:
$F(1,43) = 5.60$, $p < .03$; Workload: $F(1,43) = 65.51$, $p < .0001$).
Generally, subjects said that color inconsistency caused their
workload to increase and their SA to become poorer.

Color inconsistency also appeared to interact with groups, *Wilks'
Lambda* = 0.41, $F(24,141) = 1.71$, $p < .03$. Univariate ANOVAs showed
that the interaction was reliable in only the ratings for flash (see
Table 5), $F(1,43) = 3.86$, $p < .004$; all other p's > .1.
Interpretation of this interaction is not completely clear;
nevertheless, it appears that color inconsistency *facilitated* flash
awareness when the color task was performed alone but *interfered* with
flash awareness when the flash task was performed either alone or in
combination with the location task.

Flash probability effects. Flash probability had no main or
interaction effects on the ratings (p's > .3).

15

Table 4. *Subjective SA and Workload Ratings*

Flash Probability	Colors	Awareness of Location	Color	Flash	Workload
Equal	Consistent	.37	.29	.29	.12
Equal	Inconsistent	.18	.22	.21	.36
Red	Consistent	.30	.27	.28	.15
Red	Inconsistent	.15	.22	.23	.37

Table 5. Ratings of Flash Awareness by Group and Color Inconsistency

Group	Consistent Colors	Inconsistent Colors	Difference
L	.23	.27	-.04
C	.19	.31	-.12 ($p < .05$)
F	.34	.16	.18 ($p < .01$)
LC	.22	.28	-.06
LF	.37	.13	.24 ($p < .01$)
CF	.28	.22	.06
LCF	.34	.16	.18 ($p < .01$)

Task Performance Measures

Because which tasks subjects performed was a between-subjects variable, not all groups performed the same tasks. Therefore, the task performance data were analyzed in three separate MANOVAs, one MANOVA for all of the groups performing a common task. Thus, there was a location "group" consisting of groups L, LC, LF, and LCF, there was a color group consisting of groups C, LC, CF, and LCF, and there was a flash group consisting of groups F, LF, CF, and LCF (see footnote from Table 1 for definitions of these group labels). Note that there were four groups in each analysis, and that LCF was the only group included in all three analyses.

Location task. Table 6 shows the location errors from the four within-subject conditions. As can be seen, location error ranged from 4.5 to 5.0 degrees of visual angle (Fracker, 1991a, reported errors ranging from 2.4 to 7.9 degrees). This range should be compared to the smallest errors possible in this experiment (placing an object

just one grid node from its correct location): 1 deg vertical, 1.33 deg horizontal, and 1.67 deg diagonal. Compared to these values, the obtained errors seem rather large. Therefore, an effort was made to determine the errors that would result from pure random guessing. A short computer program was written in order to simulate 20,000 trials in which subjects randomly guessed objects' locations. This simulation produced average errors as follows: 5.2 deg vertical, 6.9 deg horizontal, and 8.7 deg diagonal.

Table 6. Location Error by Color Inconsistency and Flash Probability (degrees of visual angle)

Flash Probability	Colors	Red	Blue	Green
Equal	Consistent	4.6	4.5	4.9
Equal	Inconsistent	4.8	4.8	4.7
Red	Consistent	4.6	4.7	4.7
Red	Inconsistent	4.9	5.0	4.8

In order to determine whether subjects may simply have been guessing in the location task, the low figure of 5.2 deg was subtracted from each subjects' location error, and the resulting differences were tested in order to determine if they were different from (less than) zero. Failure to reject the null hypothesis would mean that subjects' location error data are consistent with random guessing. The results of this analysis are shown in Table 7. As can be seen, subjects who performed the location task by itself or with only the color task do not appear to have guessed at object locations. On other hand, subjects who performed the location task in combination with the flash task do appear to have been guessing.

Table 7. Comparison of Location Errors to Random Guesses.

Group	Mean[1]	SD	Minimum	Maximum	t	p[2]
L	-1.03	1.31	-3.02	1.83	-4.16	< .0002
LC	-0.52	1.33	-2.91	2.06	-2.07	< .025
LF	-0.07	1.41	-2.25	1.95	-0.26	> .35
LCF	0.01	1.36	-2.34	2.54	0.05	> .45

[1] Location error minus 5.15 degrees of visual angle
[2] One tailed test for H1: Error - 5.15 < 0.

Although groups performing the flash task were more likely to guess than others, the effect of groups on location error did not achieve statistical significance in the MANOVA ($p > .3$). Further, neither flash probability nor color inconsistency had any reliable main or interaction effects on location error (all p's $> .2$).

Color task. In the initial MANOVA, the accuracy and speed of color task responses were averaged over object colors and analyzed together. If an effect was significant, then univariate ANOVAs were undertaken to determine whether the effect was in accuracy, speed, or both. If an effect was in accuracy (for example), then further ANOVAs were performed to determine the effects on accuracy by object color.

Unlike in the location task, the color task data showed main effects of both group and color inconsistency (Group: *Wilks' Lambda* = 0.48, $F(6,46) = 3.34$, $p < .009$; Color Inconsistency: *Wilks' Lambda* = 0.16, $F(2,23) = 59.26$, $p < .0001$; all other p's $> .1$. Subsequent univariate ANOVAs showed that the group effect was significant only in the reaction time data, $F(3,24) = 6.43$, $p < .003$ (accuracy: $p > .3$) and was further reliable for all three colored objects (Red: $F(3,24) = 5.53$, $p < .005$; Blue: $F(3,24) = 5.94$, $p < .004$; Green: $F(3,24) = 5.21$, $p < .007$). In all three colors, the effect was the same: color task reaction times were significantly slowed by the addition of the location task but not the flash task (p's $< .01$; see Table 8).

Table 3. Color Probe Reaction Time by Group (milliseconds)

Object Color

Group	Red	Blue	Green
C	1006	1142	1171
LC	1715	1789	1748
CF	1370	1446	1419
LCF	1648	1984	1730

Univariate ANOVAs showed that the color inconsistency effect evident in Table 9 was reliable in both accuracy and reaction time (accuracy: $F(1,24) = 70.59$, $p < .0001$; reaction time: $F(1,24) = 34.57$, $p < .0001$). Further, for both accuracy and reaction time, the effect was reliable for all three colored objects (ACCURACY: Red: $F(1,24) = 12.05$, $p < .002$; Blue: $F(1,24) = 26.65$, $p < .0001$; Green: $F(1,24) = 48.16$, $p < .0001$; REACTION TIME: Red: $F(1,24) = 7.54$, $p < .02$; Blue: $F(1,24) = 9.86$, $p < .005$; Green: $F(1,24) = 12.34$, $p < .002$). Again, the effect was the same for all three colors: compared to color consistency, inconsistency caused responses to become both slower and less accurate (no speed-accuracy tradeoff).

Table 9. Color Inconsistency Effect on Color Task Accuracy (percent correct) and Reaction Time (milliseconds).

	Accuracy			Reaction Time		
Colors	Red	Blue	Green	Red	Blue	Green
Consistent	90	87	90	1341	1485	1404
Inconsistent	74	66	67	1528	1695	1629

Flash task. The primary performance measure for the flash task was sensitivity, the ability to discriminate flashes from non-flashes. In detection tasks, reaction time is subject to response bias as is accuracy, also called the "hit" rate. Because of this bias, reaction time and accuracy are uninterpretable; thus, these two measures were not analyzed.

An accurate measure of sensitivity could be calculated for flashes in general but not for flashes of specific colored objects (because of the inability to allocate "correct rejections" to specific colors). Therefore, a univariate ANOVA was carried out on global sensitivity with no attempt to determine how sensitivity varied across object color. Color inconsistency was found to have the only reliable effect on flash task sensitivity, $F(1,24) = 13.44$, $p < .002$; all other p's > .1. Compared to color consistency, inconsistency led to a small but statistically significant decrease in sensitivity (from .96 to .95).

DISCUSSION

The present experiment attempted to answer four questions left unresolved by Fracker's (1991a) experiments. First, was location error unaffected by color inconsistency because subject responses were less precise than the measurement process assumed? Second, was detection sensitivity unaffected by color inconsistency because color processing was not a sufficiently integral component of the detection task? Third, were location errors for red objects smaller than for other objects because more attention had been allocated to them? Fourth, did combining the location, color, and detection tasks produce performance decrements compared to performance of each task alone? Implications of the present results for these four questions are discussed in turn.

Color Inconsistency and Location Error

Color inconsistency led to slower and less accurate color probe responses, replicating Fracker's (1991a) finding that inconsistency

19

successfully incre sed the difficulty of the color identification task. The question, then, is whether this increased difficulty drew resources away from the location task thereby leading to greater location error. The answer is not clear. On one hand, consider subjects' ratings of workload and of time spent attending to object attributes. According to these ratings, color inconsistency increased mental workload which, in turn, led subjects to allocate more attention to color and less to location, just as resource theory would predict (Kahneman, 1973; Norman and Bobrow, 1975). On the other hand, these subjective ratings are not supported by the location error data; those data were unaffected by color inconsistency.

Although no solid explanation for the contradiction between performance and subjective measures can be given, their dissociation is in keeping with frequently reported dissociations of subjective and performance measures of mental workload (Yeh and Wickens, 1988). A similar example of such dissociation also appears in the subjective ratings of attribute awareness. In contrast to the location and color task performance data, these ratings indicated that color inconsistency reduced awareness of location but not of color (although the means were in the predicted direction, the effect was not statistically reliable). Thus, subjective measures of SA, like subjective measures of workload, seem to dissociate from their performance counterparts. A tempting speculation is that subjective SA ratings reflected the amount of attention allocated rather than the accuracy of information processed. Perhaps subjects assumed that if they reallocated attention away from location to color, then their awareness of color must have benefited while awareness of location must have suffered. This speculation suggests that subjective ratings of SA may reflect a process of rational deduction as much as (if not more than) a process of introspection.

While different underlying processes could explain why subjective ratings were affected by color inconsistency whereas location error was not, another explanation is that location error continues to be a poor, unreliable measure of location awareness. Two facts suggest that subjects were barely able to perform the location task. First, adding the flash task to the location task produced location responses that were indistinguishable from random guesses. Second, location responses obtained while the location task was performed alone, though better than random guesses, were hardly distinguishable from those obtained concurrently with the flash task. Further, these results were obtained in spite of efforts to improve the ease of the flash task. In fact, introduction of the grid used to constrain object movements, and of the touch screen to facilitate subject responses, seem to have had little effect on location errors: the average error of 4.8 degrees observed in the present experiment is about the same as the average of 4.7 degrees (for enemy objects) reported in Fracker (1991a). Given these observations, two alternative conclusions are possible: either people are not very good at monitoring specific

object locations, or location probes are not effective measures of what people know about those locations. More research is needed to determine which conclusion is correct.

Color Inconsistency and Detection Sensitivity

Here the present experiment was successful: requiring subjects to identify the color of flashing objects was sufficient to make detection sensitivity sensitive to color inconsistency. This conclusion is derived from the contrast of the present results with Fracker (1991a) where detection sensitivity was *not* sensitive to color inconsistency. The main differences with respect to the earlier study are two: (1) the signal here was a brief flash rather than proximity to the subject's aircraft, and (2) color identification was an explicit component of the present detection task whereas it was only implicit in the previous task. Of these two differences, it seems most likely that the second accounts for the different results. Thus, comparison of the two experiments suggests that the effect of color inconsistency on flash detection *depends* upon making color identification integral to the flash task. This outcome then argues against a common resource underlying both location and color processing. Whether this means that there are separate resources for processing these object attributes (Navon and Gopher, 1979), or alternatively that outcome conflict accounts for the interference (Navon and Miller, 1987), can not be determined from the present experiment.

In terms of the utility of detection sensitivity as a measure of SA, the present results are encouraging. Fracker (1991a) showed that detection sensitivity could pick up the effects of combat intensity on location awareness; the present results show that the same measure can respond to the effect of inconsistency on identity awareness. Taken together, these studies also underscore the need to tailor the sensitivity measure so that it will respond to the attribute of interest. This need, as Fracker (1991b) observed, suggests that detection sensitivity may be a difficult measure to implement in many situations. Nevertheless, sensitivity may often be a useful tool for assessing SA providing that some effort is devoted to its thoughtful implementation.

Attention Allocation and Location Error

Unfortunately, the present attempt to relate location error to attention allocation must be judged inconclusive. While manipulation of flash probability led subjects to report more time spent attending to red objects when their flash probability increased, this effect occurred even for those subjects who did not perform the flash task, suggesting a possible "demand effect" of the ratings procedure. Further, flash probability had no effect on flash detection sensitivity. Thus, although the subjective ratings indicate that

21

flash probability influenced attention allocation, flash task performance data do not. A possible explanation is that the manipulation of flash probability (1/3 versus 1/2 for red objects, 1/3 versus 1/6 for green objects) was not sufficient to alter subjects' attention allocation strategy enough to influence task performance. Future research might try a more dramatic manipulation. For the present, the failure of flash probability to influence location error (as well as color identification, or subjective ratings of SA) is ambiguous with respect to where the problem lies: inadequate sensitivity on the part of location error, or an inadequate manipulation of attention allocation strategy.

Dual- and Triple-Task Decrements

A major focus of the present research was on how the location, color, and flash tasks interacted with each other. Of the three, only color identification was affected by the concurrent performance of the other tasks. Further, color identification was affected by the concurrent performance of the location task but not of the flash task. In retrospect, this result is not surprising. First, the flash task required subjects to attend to color (because subjects had to indicate the color of the object that had flashed) and thus could only support color processing required in the color task. Second, as the subjective ratings indicate, when subjects performed both the location and color tasks, they had to choose between the two as to which would receive the most attention. Thus, attending to location drew attention away from color and resulted in slower responses to the color task probes.

At first glance, the implicit trade-off between attending to color or location seems inconsistent with the conclusion that color and location processing do not share a common resource (see the discussion of location error and flash detection sensitivity above). One might note, however, that both require visual attention in addition to whatever separate central resources that they may need. Further, attending to color requires attending only to the object itself; on the other hand, attending to location requires attending to both the object and the surrounding environment. Thus, attending to the surrounding environment seems to reduce the amount of attention given to the object, and allocating more attention to the object appears to reduce the amount paid to the environment. This tradeoff, then, seems to reflect the limitations of attention allocation in the visual field identified by Eriksen and Yeh (1985): attending to a larger visual space reduces the quality of perceived information.

Conclusions

Identity probe and detection task performance both appear to be good measures of SA. Both responded well to color inconsistency in the present experiment; in Fracker (1991a), both also responded well

to combat intensity. On the other hand, in spite of efforts to
improve it, location error remains a measure of questionable
usefulness. While future research may be able to pinpoint the source
of difficulty, the present experiment offers no hope that location
error can be improved as measure of SA.

Subjective measures of attention allocation, SA, and workload all
proved useful in the present experiment. As is often the case in
applied settings, these measures were more responsive to the
experimental manipulations than were any of the more objective
performance measures (cf., Hughes, Hassoun, Ward, and Rueb, 1990; Ward
and Hassoun, 1990). This responsiveness, combined with their ease of
use, may make them attractive to many users. But the present
experiment offers a caution: subjective measures may be measuring
subjects' rational inferences about their SA or workload rather than
the results of their introspections. Until researchers have a better
understanding of just how people produce their responses to subjective
rating scles, caution in their use would seem to be in order.

REFERENCES

Budescu, D. V., Zwick, R., and Rapoport, A. (1986). A comparison of
the eigenvalue method and the geometric mean procedure for ratio
scaling. *Applied Psychological Measurement, 10*, 69-78.

Endsley, M. R. (1989). A methodology for the objective measurement
of pilot situation awareness. In *Proceedings of the NATO AGARD
Conference on Situational Awareness in Aerospace Operations* (AGARD-
CP-478). Springfield, VA: National Technical Information Service.

Eriksen, C. W., and Yeh, Y. (1985). Attention allocation in the
visual field. *Journal of Experimental Psychology: Human Perception
and Performance, 11*, 583-597.

Fracker, M. L. (1991a). *Measures of situation awareness: An
experimental evaluation.* Technical Report in preparation. Wright-
Patterson AFB OH: Armstrong Laboratory, Crew Systems Directorate.

Fracker, M. L. (1991b). *Measures of situation awareness: Review and
future directions.* Technical Report in preparation. Wright-
Patterson AFB OH: Armstrong Laboratory, Crew Systems Directorate.

Fracker, M. L., and Davis, S. A. (1990). Measuring operator
situation awareness and mental workload. In *Proceedings of the
Fifth Mid-Central Ergonomics/Human Factors Conference.* Dayton, OH:
University of Dayton.

Hirst, W., & Kalmar, D. (1987). Characterizing attentional
resources. *Journal of Experimental Psychology: General, 116*, 68-
81.

Hughes, E. R., Hassoun, J. A., Ward, G. F., and Rueb, J. D. (1990). *An assessment of selected workload and situation awareness metrics in a part- mission simulation: Final report* (ASD-TR-90-5009). Wright- Patterson AFB OH: Aeronautical Systems Division.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall.

Keppel, G. (1982). *Design and analysis: A researchers handbook.* Englewood Cliffs, NJ: Prentice-Hall.

Macmillan, N. A., and Creelman, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and "non-parametric" indexes. *Psychological Bulletin, 107,* 401-413.

Marshak, W. P., Kuperman, G., Ramsey, E. G., and Wilson, D. (1987). Situation awareness in map displays. In *Proceedings of the Human Factors Society 31st annual meeting* (pp. 533-538). Santa Monica, CA: Human Factors Society.

Navon, D. (1984). Resources--A theoretical soupstone? *Pyschological Review, 91,* 216-234.

Navon, D. (1985). *Do people allocate limited resources among concurrent activities?* (Technical Report No. IPDM 26, April). Haifa, Israel: University of Haifa, Laboratory for Information Processing and Decision Making.

Navon, D. (1990). Exploring two methods for estimating performance tradeoff. *Bulletin of the Psychonomic Society, 28,* 155-157.

Navon, D., and Gopher, D. (1979). On the economy of the human processing system. *Psychological Review, 86,* 214-255.

Navon, D., and Miller, J. (1987). Role of outcome conflict in dual-task interference. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 435-448.

Norman, D. A., and Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7,* 44-64.

Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology, 15,* 234-281.

Vidulich, M. A. (1989). The use of judgment matrices in subjective workload assessment: The Subjective WORkload Dominance (SWORD) technique. In *Proceedings of the Human Factors Society 33rd annual meeting* (pp. 1406- 1410).

Ward, G. F., and Hassoun, J. A. (1990). *The effects of head-up display pitch ladder articulation, pitch number location and horizonline length on unusual attitude recoveries for the F-16* (ASD-TR-90-5003). Wright- Patterson AFB OH: Aeronautical Systems Division.

Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson & R. Pew (Eds.), *Attention and performance VIII* (pp. 239- 258). Hillsdale, NJ: Erlbaum.

Wickens, C. D. (1984). *Engineering psychology and human performance.* Columbus, OH: Merril.

Yeh, Y., and Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. *Human Factors, 30,* 111-120.